




 viiiiiiiiiiii / Flatiron\_Phase1\_Final\_Vi\_Bui

Phase 1 Project - Jupyter Notebook and Presentation

☆ 0 stars    🍴 0 forks

 Star Unwatch ▾

&lt;&gt; Code

 Issues Pull requests Actions Projects Wiki Security main ▾

...



viiiiiiiiiii Add files via upload ...

now ⌚ 14

[View code](#) README.md 

# Flatiron Data Science - Phase 1 Final Project - Vi Bui

- Student Name: Vi Bui

## Microsoft Studios - Building a Movie Studio for the Future

### Project Overview

Microsoft would like to create a movie studio. They've asked our company to explore what types of films are currently doing the best at the box office.

**Data, Methodology, and Analysis:** we've explored data from Rotten Tomatoes, Box Office Mojo, IMDB, TheMovieDB, and The Numbers. After much research, we've decided to build a strong foundation for Microsoft Studios by analyzing Production Budget, Worldwide Gross, Genre, Studio, Release Date, and Box Office Revenue data.

**Results & Recommendations:** After analyzing data from databases containing movie data ranging from 1915-2019, 5234 movies, 99 genres, and 172 Studios, and later narrowing our focus to the past 10-20 years and the top 5 studios, we have recommendations on initial budget considerations, launch timing, and genre assortment.

## Business Objectives

---

Make data-driven recommendations to set Microsoft Studios up for success:

1. Determine baseline production budget
2. Understand competitive and benchmark landscape
3. Propose launch timing
4. Show Genre Efficiency to make informed decisions on genre assortment

## Data Understanding and Exploration

---

After exploring numerous movie databases, Box Office Mojo, Rotten Tomatoes, and The Numbers will be our main data sources for Phase 1 analysis, while other data will be used for future analysis.

### Data explored:

\*\* denotes data used in current (Phase 1) analysis

^ denotes data for future analysis

1. Box Office Mojo - Gross: bom.movie\_gross.csv\*\*
2. IMDB - Name Basics: imdb.name.basics.csv ^
3. IMDB - Name Akas: imdb.title.akas.csv.gz
4. IMDB - Title Basics: imdb.title.basics.csv.gz
5. IMDB - Title Crew: imdb.title.crew.csv.gz
6. IMDB - Title Principles: imdb.title.principals.csv.gz
7. IMDB - Title Ratings: imdb.title.ratings.csv.gz ^
8. Rotten Tomatoes - Movie Info: rt.movie\_info.tsv \*\*
9. Rotten Tomatoes - Reviews: rt.reviews.tsv ^
10. The Movie Database - Movies: tmdb.movies.csv ^
11. The Numbers - Movie\_Budgets: tn.movie\_budgets.csv \*\*

**Merged The Numbers (Movie, Release Date, Production Budget and Worldwide Gross) data and Box Office Mojo (Movie and Studio) data to analyze:**

1. Correlation between Production Budget and Worldwide Gross for All Movies \*and Top Five Grossing Studios
2. Seasonality (monthly trends) by year for movies released between 2009-2019
3. Competitive landscape

Created Studio & Genre dataset (using Rotten Tomatoes Movie Info) to analyze:

3. Genre Efficiency

## Data Exploration, Cleansing, and Preparation

---

### Data Exploration

Outlined in detail in comments: files explored, how data was chosen, which files will be used for Phase 1 proposal, and which files will be used in future analysis

### Data Cleansing

Dropped duplicates, NaN values, and unnecessary columns; continuously cleansed data as necessary

### Data Preparation

Core variables: Production Budget, Worldwide Gross, Genre, Studio, Release Date, Box Office Revenue. Chose data that best served analysis, merged data, and created clean datasets and visualizations for analysis

## Business Objective #1 - Baseline Budget

---

- **Objective:** Determine baseline production budget
- **Results:**
  1. Data shows strong correlation (0.75) between Production Budget and Worldwide Gross for **all movies** in the database (5782 films from 1915-2019). Average Production Budget is 34,033,480 and Average Worldwide Gross is 100,761,506.
  2. Data shows the strongest correlation between the three databases observed (0.79) between Production Budget and Worldwide Gross for **all films from the past ten years** (1992 films from 1998-2018). Average Production Budget is 42,638,335 and Average Worldwide Gross is 134,678,815.
  3. Data shows the weakest, though still strong, correlation (0.72) for Production Budget and Worldwide Gross for **movies from the top fives studios in the last 10**

**years** (465 films from 1998–2018). Average Production Budget is 79,098,925 and Average Worldwide Gross is 272,398,792.

- **Recommendations:** Plan for Baseline Production Budget: 43MM, Midrange Budget: 61MM, and High-End/Blockbuster Budget: 79MM.
- **Source:** The Numbers Database

Data shows strong correlation (0.75) between Production Budget and Worldwide Gross for all movies in the database (5782 films from 1915–2019). Average Production Budget is 34,033,480 and Average Worldwide Gross is 100,761,506.



## Business Objective #2 - Competitive & Benchmark Landscape

---

- **Objective:** Understand competitive and benchmark landscape
- **Results:** The Top Five Studios in the Box Office Mojo are: The Walt Disney Company ('BV'), Universal Pictures ('Uni.'), 20th Century Fox Studios ('Fox'), Warner Bros. ('WB'), and Sony Studios ('Sony')
- Data shows relatively strong (0.72) correlation between Production Budget and Worldwide Gross for the top five studios; though slightly lower than overall movie data
- **Recommendations:** We will use these studios as the benchmark landscape for our Phase 1 analysis, but strongly suggest we analyze "Studio Efficiency" metrics to determine whether Microsoft would like to take a different genre assortment and launch approach to set themselves apart from the top five studios
- **Source:** 1. The Numbers and 2. Box Office Mojo Databases

Part 2: we would like to analyze worldwide gross as it relates to movie *studio*

Top 5 Studios - Worldwide Gross from 2009–2019



Top 5 Studios - Production Budget from 2009–2019



## Business Objective #3 - Launch Timing

---

- **Objective:** Propose launch timing

- **Results:** Movies are released and gross the highest in Summer (May-July) and Holiday (Nov-Dec) months; this trend is more pronounced when looking at the top five studios
- **Recommendations:** While there is appeal in launching during the most popular months, we think there is an opportunity to capture "off" months while other studios are less active
- **Source:** 1. The Numbers and 2. Box Office Mojo Databases

#### Seasonality Trends - Top 5 Studios (2009-2019)



- Business objective: explore Seasonality
- We now have all the data needed to work on Business Question #2: Seasonality and Launch Timing

#### Seasonality Trends - All Movies (2009-2019)



- From above, observe the difference vs. top five studio seasonality
- QUICK OBSERVATION: the top five studios have strong worldwide gross in February, June, and November

#### Monthly Trends by Year - All Movies (2009-2019)



#### Monthly Trends by Year - Top 5 Studios (2009-2019)



## Business Objective #4 - Genre Assortment

---

- **Objective:** Show "Genre Efficiency" to make informed decisions on genre assortment
- **Results:** Comedy has highest box office sales in the last 20 years with 133 films; Drama is the most "categorized" genre (highest number of films); Science Fiction is the "most efficient" genre with the highest sales per movie; Action & Adventure are the next most efficient genre
- **Recommendations:** We recommend a healthy assortment of films in "mainstream" genres (Comedy, Drama), "efficient" genres (Science Fiction, Romance), and of

course, the hybrid: Action & Adventure! As a new studio, we would again propose a different approach (as was our launch timing proposal) of launching with Science Fiction, Romance, or Action & Adventure.

- **Source:** Rotten Tomatoes

Genre Efficiency - Worldwide Sales Per Film

Top Left Quadrant: Highly Efficient

Bottom Right: Mainstream/Popular



## Evaluation and Conclusions

---

We hope our preliminary analysis gives Microsoft Studios a solid foundation to understand the movie business. Our overall objective was to make Microsoft understand the fundamentals of the landscape (production budget, worldwide gross, competitive landscape/studios) and propose a launch plan both anchored in data and outside the box (launch timing, genre assortment)

### Summary of recommendations

- Plan for Production Budget of \$30MM per film, particularly to be competitive with major studios
- Consider launching outside of competitive months either in Q1 (Jan-Mar) or Q3 (Sep)
- Launch with a mixture of "mainstream" genres (Drama, Comedy) and "efficient" genres (Science Fiction, Romance)
- With the information provided, work with Microsoft to determine Microsoft's Studios goals (studio brand identity, short and long term fiscal goals, etc.)

### Further considerations:

- Depending on how Microsoft would like to move forward, there are several areas of opportunity to dig deeper. We would like to dig deeper into launch timing - for instance, understanding studio production budget and gross through the years (early years to more mature years)
- Determine when studios reach a point of "maturity"

- Note: we did not perform a "return on investment" analysis because there are a lot of factors we'd like to consider outside of Worldwide Gross-Production Budget
- Other areas of opportunity: gain deeper understanding and learn from efficient and/or esoteric studios and genres

## Future Work

---

This is just the beginning! We can't wait to work together to launch Microsoft Studios with a bang, and build a studio for the future.

### Future work:

- Determine how/if ratings impact performance
- Model future trends for genres -> Discover/create "new" genres
- Build long term movie launch strategy (type/genre by month)
- Explore other movie consumption formats (streaming)
- Analyze Franchise movie data

## Releases

No releases published

[Create a new release](#)

## Packages

No packages published

[Publish your first package](#)

## Languages

● **Jupyter Notebook** 100.0%